Haute disponibilite logicielle et BlueMind

Cette page n'est plus actualisée. À partir de BlueMind 4.8, veuillez consulter la **nouvelle documentation BlueMind**

Présentation

II est possible de mettre en place un système de haute disponibilité logicielle (HA pour High Availability en anglais) s'intégrant avec BlueMind.

Le présent document donne les recommandations et informations sur le système BlueMind nécessaire pour pouvoir intégrer la solution de messagerie dans une infrastructure de haute disponibilité.

Sur cette page :

- Présentation
- Préparation du système
  - Espace de stockage
  - Réseau
  - Scripts de supervision
- Configuration de la Haute disponibilité
  - Données et services à gérer
  - STONITH

Les solutions logicielles tierces mentionnées dans le présent document sont données à titre d'exemple. Cette liste ne saurait être exhaustive.

Préparation du système

Note : les deux serveurs en jeu doivent respecter les recommandations de dimensionnement matériel définies dans le document suivant : Dimensionnement matériel

Espace de stockage

Le contenu à partager entre les deux serveurs peut l'être soit sur un espace de stockage partagé comme par exemple un SAN (*Storage Area Network*) , soit via une réplication de données entre deux espaces de stockages séparés.

La haute disponibilité via un mécanisme de réplication peut induire des problèmes majeurs d'accès aux ressources disques partagées qui surviennent le cas échéant dans des cas de pertes de services. Le cas le plus courant de soucis d'accès aux ressources ayant un impact potentiellement désastreux est le scénario de *split-brain* .

Le composant *cyrus-imap* ne supporte pas les stockages de type NFS pour la gestion des méta-données. Quel que soit le choix retenu pour le type de stockage répliqué, il faut donc un stockage de type *block-device* se basant par exemple sur les technologies Fibre Channel ou iSCSI pour le répertoire `/var/spool/cyrus/meta`.
Tous les autres répertoires comme `/var/spool/cyrus/data` et `/var/lib/cyrus` peuvent quant à eux être stockés sur des espaces de stockages montés en NFS.

Données à rendre disponible entre les deux serveurs

Les données situées dans les répertoires suivants sont celles qui doivent être visibles par les deux serveurs et dont l'accès doit être géré par le système de gestion de la HA :

- /var/spool/bm-docs
- /var/spool/bm-elasticsearch
- /var/spool/bm-hsm
- /var/spool/cyrus
- /var/spool/postfix
- /var/spool/sieve
- /var/spool/bm-hollowed
- /var/spool/bm-mapi

À ces derniers doit être ajoutée la base de données cyrus située dans le répertoire suivant :

- `/var/lib/cyrus`
- `/var/lib/postgresql`

Ces données doivent donc se trouver sur un espace de stockage permettant au serveur passif d'accéder aux données en cas de bascule, par exemple un stockage SAN, un cluster GFS, ou autre..

RAPPEL : `/var/spool/cyrus/meta` ne doit en aucun cas être stocké sur un montage NFS, en revanche `/var/spool/cyrus/data` peut l'être

Réseau

Afin de fonctionner correctement, BlueMind doit être accessible via une seule URL/IP, il est donc recommandé d'utiliser un système pouvant gérer des adresses IP flottantes (ou virtuelles).

L'URL d'accès sur les frontend BlueMind doit obligatoirement être toujours la même.

Scripts de supervision

Voir la page dédiée Supervision

Configuration de la Haute disponibilité

Sans STONITH (voir ci-après), il ne faut pas activer la bascule automatique au risque d'avoir des défaillances *split-brain* et des corruptions de données (voir encart dans le paragraphe dédié) qui ne seront pas prises en compte par le support BlueMind.

Données et services à gérer

Configuration de BlueMind à synchroniser

Les fichiers de configurations BlueMind à synchroniser en temps réel par le système de gestion de la HA sont situés dans le répertoire /etc.

Il faut également synchroniser les fichiers :

- /usr/share/bm-elasticsearch/config/elasticsearch.yml
- /etc/aliases
- /etc/aliases.db
- /etc/sysctl.conf
- /etc/ssl/certs/bm_cert.pem
- /var/lib/bm-ca/ca-cert.pem

Pour réaliser une synchronisation en temps réel des fichiers de configuration, il est possible de se baser sur les exemples suivant :

- incron, basé sur inotify, permet de lancer des tâches en fonction de l'état d'un fichier par exemple. La documentation officielle est disponible sur le site de l'éditeur.
- les fichiers peuvent être copier par *rsync over ssh* par exemple, comme présenté sur ce site.
- d'autres outils existent comme l syncd et csync2

Gestion de la mise à jour de BlueMind

Les grandes étapes de la mise à jour d'un déploiement en Haute Disponibilité de BlueMind sont décrites ci-après :

- Avant de lancer la mise à jour de BlueMind, désactiver les services de gestion de la haute disponibilité.
- Mettre à jour les paquets sur les deux serveurs.
- Puis **sur le serveur principal uniquement** possédant l'adresse IP publique, réaliser la configuration post-installation comme indiqué au paragraphe : Configuration post-installation.

STONITH

STONITH, pour *Shoot The Other Node In The Head*, est une technique de *fencing*, ou isolement, dans la gestion de clusters. Le principe est de pouvoir éteindre le serveur défaillant d'un cluster à distance, soit logiciellement, soit directement en lui coupant son alimentation électrique.

Ce type de fonctionnement se situe au niveau de l'infrastructure matérielle.

Cette sécurité permet de diminuer fortement les risques de corruption de données dans des cas de pertes de service complexes, par exemple comme dans le cas d'une défaillance de type *split-brain* qui va conduire les deux serveurs à se considérer unique maître et tenter d'accéder en même temps à la ressource de stockage partagée. Dans le cas d'une haute-disponibilité basée sur une réplication de données, les risques de corruption de données sont importants.

Cette technique peut par exemple être mise en place avec des outils IPMI (*Intelligent Platform Management Interface*). IPMI est une spécification d'interfaces de gestion de machines, mais il est possible d'en trouver des implémentations, comme par exemple freeIPMI, OpenIPMI, ipmitool, ...

L'implémentation côté matériel peut se faire par un matériel dédié ou simplement par l'utilisation par exemple des cartes iDRAC pour du matériel du constructeur DELL.